BMJ Open

Development of a computable phenotype to identify a transgender sample for health research purposes: a feasibility study in a large linked provincial healthcare administrative cohort in British Columbia, Canada

Ashleigh J Rich ,[1,2] Tonia Poteat,[3] Mieke Koehoorn,[1] Jenny Li,[2] Monica Ye,[2] Paul Sereda,[2] Travis Salway ,[4] Robert Hogg[2,4]

To cite: Rich AJ, Poteat T, Koehoorn M, *et al*. Development of a computable phenotype to identify a transgender sample for health research purposes: a feasibility study in a large linked provincial healthcare administrative cohort in British Columbia, Canada. *BMJ Open* 2021;**11**:e040928. doi:10.1136/bmjopen-2020-040928

► Prepublication history and supplemental material for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-040928).

Received 26 May 2020
Revised 03 March 2021
Accepted 05 March 2021

Check for updates

© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Ashleigh J Rich;
ajrich@mail.ubc.ca

ABSTRACT

Objectives Innovative methods are needed for identification of transgender people in administrative records for health research purposes. This study investigated the feasibility of using transgender-specific healthcare utilisation in a Canadian population-based health records database to develop a computable phenotype (CP) and identify the proportion of transgender people within the HIV-positive population as a public health priority.

Design The Comparative Outcomes and Service Utilization Trends (COAST) Study cohort comprises a data linkage between two provincial data sources: The British Columbia (BC) Centre for Excellence in HIV/AIDS Drug Treatment Program, which coordinates HIV treatment dispensation across BC and Population Data BC, a provincial data repository holding individual, longitudinal data for all BC residents (1996–2013).

Setting British Columbia, Canada.

Participants COAST participants include 13 907 BC residents living with HIV (≥19 years of age) and a 10% random sample comparison group of the HIV-negative general population (514 952 individuals).

Primary and secondary outcome measures Healthcare records were used to identify transgender people via a CP algorithm (diagnosis codes+androgen blocker/hormone prescriptions), to examine related diagnoses and prescription concordance and to validate the CP using an independent provider-reported transgender status measure. Demographics and chronic illness burden were also characterised for the transgender sample.

Results The best-performing CP identified 137 HIV-negative and 51 HIV-positive transgender people (total 188). In validity analyses, the best-performing CP had low sensitivity (27.5%, 95% CI: 17.8% to 39.8%), high specificity (99.8%, 95% CI: 99.6% to 99.8%), low agreement using Kappa statistics (0.3, 95% CI: 0.2 to 0.5) and moderate positive predictive value (43.2%, 95% CI: 28.7% to 58.9%). There was high concordance between exogenous sex hormone use and transgender-specific diagnoses.

Conclusions The development of a validated CP opens up new opportunities for identifying transgender people for inclusion in population-based health research using administrative health data, and offers the potential for much-needed and heretofore unavailable evidence on health status, including HIV status, and the healthcare use and needs of transgender people.

Strengths and limitations of this study

- ► This study demonstrates the feasibility of developing and validating a computable phenotype (CP) for identification of a transgender sample, using a population-based representative source population and healthcare records.
- ► A major contribution of this study is the ascertainment of the population of transgender people living with HIV in the Canadian province of British Columbia (BC), in a universal healthcare setting, using a CP and capacity to estimate the prevalence of transgender status among the population living with HIV in the province.
- ► Development of a validated CP algorithm using diagnosis and prescription data to identify transgender samples in administrative data without other gender identity ascertainment measures lays the foundation for future investigation of transgender-specific research questions related to general and HIV-specific healthcare use and health outcomes for this key population.
- ► While administrative data is an invaluable resource for answering important health and healthcare utilisation research questions, this study is limited to those transgender people accessing medical transition care in BC and may not represent the transgender population as a whole.

INTRODUCTION

Limited data on transgender people

Transgender people are often overlooked within epidemiological research and

BMJ Open: first published as 10.1136/bmjopen-2020-040928 on 25 March 2021. Downloaded from http://bmjopen.bmj.com/ on November 7, 2023 by guest. Protected by copyright.

population health surveillance due to small sample size, limited research designs and other institutional and methodological erasures.[1–3] A 2017 review of Medline-indexed literature from 1950 to 2016 found 2405 published articles including transgender people, with almost half published in the last decade.[4] A 2008 US-based meta-analysis of HIV prevalence among transgender populations found 24 studies of transgender women, and 5 additional studies of transgender men,[5] though an updated review found 43 primary studies on transgender women and 15 on transgender men published between 2006 and 2017.[6] Despite this recent increase in transgender health research in general and for HIV specifically, much of the literature has focused on transgender-specific care, mental health and HIV/sexual health,[7 8] leaving the population understudied, in particular in the broader areas of physical health and healthcare utilisation.

The erasures or exclusions of transgender persons in health studies may be explained, in part, by methodological challenges. Specific to electronic health record (EHR) data, a 2017 report identified only one transgender person among 38 5820 cancer cases in a Minnesota cancer registry,[9] clearly an undercount given that 0.4% of the US general population and 0.6% (95% credibility intervals: 0.5%–0.7%) of the Minnesota population is estimated to be transgender.[10 11] This highlights the need for improved gender ascertainment and transgender inclusion in research relying on patient records and administrative data. The establishment of best practices for measuring transgender status in research survey, such as the two-step method (measuring sex assigned at birth and current gender identity), points to a way forward for transgender-inclusive population health research.[12 13] However, innovative research methods are needed to identify transgender people in studies that rely on existing data sources (in particular EHR) and that optimise the use of transgender respondents' data in non-transgender-specific research.

Computable phenotypes (CPs) for transgender health research

Previous research in transgender health largely comprises cross-sectional studies, case reports and qualitative or observational research.[7] Much consists of clinic-based or venue-based convenience samples or lack comparison groups.[7 8] The literature is further characterised by inconsistent transgender status measurement,[14] small sample sizes and focus on the USA.[8] In response, researchers have called for advancing transgender health research methods—namely ascertainment of high-quality samples via systematic approaches—including for general population-based and health systems-based studies.[15] One opportunity for the advancement of transgender health research methods is the emerging use of CPs[16] or case ascertainment algorithms, to identify transgender samples in healthcare utilisation data. A CP is an algorithm for identifying a clinical feature, condition or set of characteristics that can be determined directly from EHR and other ancillary healthcare data systems (eg, disease registries, insurance claims data) data.[17] CPs are developed using a combination of data elements (eg, sociodemographic variables, clinical diagnoses) and value sets (ie, the selection of a set of relevant values for each data element). Development of CPs using standardised methods and definitions enables identification and inclusion of transgender persons in research, as well as replication of analyses across data sources, healthcare organisations/sites and studies. CPs have application in clinical care, surveillance and health research.

Recently, CP and other EHR-based algorithm methods have been applied in a number of settings primarily in the USA to identify transgender samples for health research.[14] Specifically, the Study of Transition, Outcomes & Gender (STRONG) study identified a transgender cohort (n=6456) using EHR data from Kaiser Permanente health plan members in California and Georgia, for investigation of general and transgender-specific health outcomes.[18] Blosnich *et al*[19] identified 3177 people with a 'gender identity disorder' diagnosis among military veterans accessing care through the US Veterans Health Administration healthcare system, for examination of mental health and other outcomes. Researchers with the US Centers for Medicare & Medicaid Services identified 4098 transgender beneficiaries using national Medicare claims data,[20] and researchers at Vanderbilt University identified 234 transgender patients in their university clinic EHR data.[16] While these cohorts represent important opportunities for advancement of transgender health research, these methods have yet to be applied widely outside the US context. This is particularly important as different jurisdictions may vary in medical billing and coding practices, healthcare system patient populations and representativeness of the general population. Specifically, in Canada, healthcare is delivered through a provincially administered universal healthcare system. As such, research using EHR provides an opportunity to develop methods for population-based, representative estimates of transgender populations within the Canadian context. Coupled with the current absence of gender ascertainment measures in population-based routinely collected data (eg, census, national government health surveys) in Canada and many other jurisdictions, this remains an evidence need.

Summary of study rationale

This study investigated the application of emerging transgender health research methods, specifically CPs, in a Canadian context for the first time, testing the feasibility of identification of a transgender sample using EHR data from a provincial healthcare administrative data-linked cohort.

METHODS

Data sources and participants

The Comparative Outcomes and Service Utilization Trends (COAST) Study

COAST is a population-based cohort study focused on health services utilisation research questions among all

BMJ Open: first published as 10.1136/bmjopen-2020-040928 on 25 March 2021. Downloaded from http://bmjopen.bmj.com/ on November 7, 2023 by guest. Protected by copyright.

people known to be living with HIV (PLWH) in the province of British Columbia (BC) and a 10% random sample comparison group of the HIV-negative general population.[21] The COAST cohort comprises individual-level longitudinal data from PLWH who have ever accessed HIV treatment in BC between 1996 and 2013, provided by Population Data BC (PopDataBC)[21] via data linkage between two provincial data sources, by personal health number: the Drug Treatment Program (DTP)[22] and the Ministry of Health. PopDataBC provides infrastructure for access to, and linkage of, longitudinal and individual-level administrative health data for all BC residents.[23] The HIV-negative general population cohort was drawn randomly from the Ministry of Health registry data by PopDataBC. The COAST study has received approval from the University of British Columbia/Providence Health Care Research Ethics Board (#H09-02905) and Simon Fraser University Office of Research Ethics (#2013 s0566). The study complies with the BC Freedom of Information and Protection of Privacy Act and did not require informed consent as it is conducted using retrospective administrative and anonymised data for research and statistical purposes only. No patients or public were involved in this study.

Drug Treatment Program

In BC, antiretroviral therapy (ART) is provided to PLWH at no cost to the patient, and distributed through the DTP.[22] The DTP contributed a provider-reported measure of transgender status for COAST.

Ministry of Health

Ministry of Health data available via COAST included insured medical service billing records for outpatient visits,[24 25] hospital (in-patient) visits,[26] prescription medications[27 28] and vital statistics.[29]

Measures and analyses

Transgender CPs

Identification of transgender cases was tested in COAST using International Classification of Disease (ICD) codes (9th and 10th editions) and exogenous sex hormone prescription use. Transgender-specific diagnoses in medical and hospital billing records included the ICD-9 codes 302.5 Trans-sexualism with unspecified history, 302.51 Trans-sexualism with asexual history, 302.52 Trans-sexualism with homosexual history, 302.53 Trans-sexualism with heterosexual history, 302.6 Gender Identity Disorder in children, 302.85 Gender Identity Disorder in adolescents or adults; and ICD-10 codes F64.0 Gender Identity Disorder of childhood, F64.2 Gender Identity Disorder of childhood, F64.8 Other Gender Identity Disorder and F64.9 Gender Identity Disorder unspecified. The full list of androgen blockers and exogenous sex hormone prescriptions included in analyses is available in the online supplemental material.

Concordance

To assess face validity and utility of diagnosis and prescription data over time in CP development (ie, whether the identified transgender sample had exogenous sex hormone prescription use and other diagnoses patterns consistent with that of transgender populations in other studies), concordance analyses evaluated the presence of at least one included diagnosis and prescription during the COAST study follow-up period with the presence of at least one included diagnosis and prescription in the last study year. Concordance was assessed between transgender-specific diagnoses, exogenous sex hormone and androgen blocker prescriptions and non-transgender specific diagnoses (ICD-9 259.9 Unspecified Endocrine Disorder and ICD-10 E34.9 Endocrine Disorder, Unspecified (see online supplemental material)). Endocrine disorder diagnosis codes are sometimes preferred by medical providers treating transgender people in response to historic exclusions of transgender-specific care from insurance coverage and to combat the stigma of transgender-specific diagnosis codes that have historically been classified as psychiatric disorders in the Diagnostic and Statistical Manual of Mental Disorders.[30] Exogenous sex hormone use, while common in transgender populations,[5 31] is not transgender-specific. Cisgender populations also use androgen blocker and sex hormone prescriptions (eg, oestrogen to treat menopausal symptoms in cisgender women, spironolactone is used for hypertension), thus exogenous sex hormone and androgen blocker prescription use cannot independently identify transgender people. At the same time, not all transgender people use hormones and some access via non-medical sources.[32 33]

Validation

In BC, transgender status is collected in the DTP via a provider-reported sex variable ('Male', 'Male to Female', 'Female to Male' or 'Female'). Patients reported as either 'Male to Female' or 'Female to Male' were classified as transgender. The provider-reported transgender measure, available for the HIV-positive cohort only, was used as a 'gold standard' for CP validation. Sensitivity, specificity, positive predictive value and kappa statistics with corresponding 95% CIs were calculated for identifying transgender people via the CPs, in the HIV-positive cohort only. Follow-up time (mean and range) for each CP group was also produced.

Demographics and chronic conditions

To further assess face validity of the transgender CP for future health research, descriptive statistics were calculated for the transgender sample produced via application of the best-performing CP from the validation analysis to both the COAST HIV-positive and HIV-negative cohorts. Descriptive statistics included COAST study key sociodemographic and health variables, specifically laboratory confirmed HIV serostatus (HIV-positive/HIV-negative), baseline age, patient's Health Authority (five provincial regions for the administration of health services that include large urban centres, suburban regions and rural/remote areas), and chronic illness burden based on standardised case definitions from the BC Ministry of Health[34] and the BC Cancer Agency.[35]

BMJ Open: first published as 10.1136/bmjopen-2020-040928 on 25 March 2021. Downloaded from http://bmjopen.bmj.com/ on November 7, 2023 by guest. Protected by copyright.

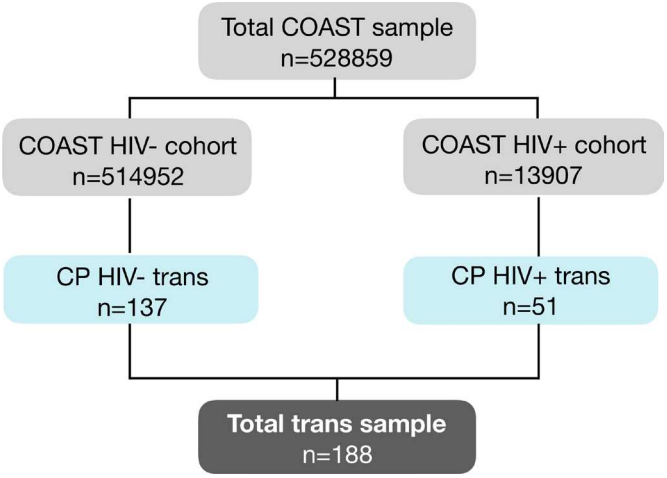

Figure 1 Total transgender sample identified using a computable phenotype with electronic health records. COAST, Comparative Outcomes and Service Utilization Trends; CP, computable phenotype.

RESULTS

The total COAST cohort included 528 859 people, of which 514 952 were HIV-negative (10% general population random sample) and 13 907 were PLWH (figure 1).

Concordance

Of the 237 people who had ever had a transgender-specific diagnosis during the study period, 19.4% also had a recent diagnosis in the last follow-up year (table 1). None had an unspecified endocrine disorder diagnosis at any time; thus, this diagnosis was excluded from all CPs. Of the 237, 79.3% had an exogenous sex hormone or androgen blocker prescription at least once during the study period and 46.4% had one in the last year.

Validation

While no one performed CP consistently well across all validation metrics, the CP with the best overall performance across test statistics was based on having received at least one transgender-specific diagnosis and at least one androgen blocker/exogenous sex hormone prescription over the study follow-up period (table 2). This CP had high specificity (99.8%, 95% CI: 99.6% to 99.8%), low sensitivity (27.5%, 95% CI: 17.8% to 39.8%), low to moderate Kappa coefficients (0.3, 95% CI: 0.2 to 0.5) and moderate positive predictive values (43.2%, 95% CI: 28.7% to 58.9%). This CP also had the second longest mean follow-up time (mean: 136.3, range: 21.0–203.0), overall similar to the other CP groups (mean: 136.5, range: 21.0–203.0; mean: 117.1, range: 24.0–198.0; mean: 130.4, range: 69.0–198.0; respectively).

Table 1 Concordance analyses for diagnoses and hormone measures

	N	%
≥Transgender ICD (ever)	237	100
≥Transgender ICD (recent)	46	19.4
Unspecified endocrine disorder use (ever)	0	0.0
Unspecified endocrine disorder use (recent)	0	0.0
≥Hormone/blocker use (ever)	188	79.3
≥Hormone/blocker use (recent)	110	46.4

ICD, International Classification of Disease.

Transgender phenotype

Applying the best-performing CP, 137 HIV-negative people and 51 HIV-positive people (188 total) were identified as transgender in the respective COAST cohorts (figure 1).

Demographics and chronic conditions

Demographic characteristics and chronic conditions for the 188 transgender people identified via the best-performing CP are presented in figures 2–4. Transgender people were geographically located throughout BC health regions. The Vancouver Coastal Health Authority region, which includes the largest municipal area in BC, had the highest concentration of transgender people (44.2%) while the Northern Health Authority region—a predominantly rural and remote area of the province—had the lowest (1.6%).[36] The HIV-positive group had a higher median age than the HIV-negative group (35 (Q1, Q3: 30, 42) and 30 (Q1, Q3: 19, 42), respectively). For the HIV-negative sample, the largest proportion of transgender people were aged 19–29 years (44.5%) and the smallest proportion aged 55 years and older (<3%). For the HIV-positive sample, the largest proportion were aged 30–34 years (25.5%) and the smallest proportion aged 55 years and older (<2%).

Overall, HIV-positive transgender people had a higher prevalence of at least one chronic condition (other than HIV) compared with HIV-negative transgender people (88.2% vs 85.4%, respectively) and of two or more chronic conditions (76.5% vs 52.6%, respectively). Specific chronic disease differences between transgender people living with and without HIV were most notable for a higher prevalence among the HIV-positive cohort of cardiovascular disease, chronic kidney disease, osteoarthritis, schizophrenia and personality disorders and chronic liver disease, but a lower prevalence for hypertension.

DISCUSSION

This study demonstrates the feasibility of identification of a sample of transgender people in a large linked provincial healthcare administrative database, using a CP based on prescriptions and diagnoses. Among a growing number of studies using EHR and CP methods to identify transgender samples for health research purposes, this is the first to do so in Canada to independently validate the CP using a 'gold standard' of provider-reported transgender status and the only to use population-based data.

BMJ Open: first published as 10.1136/bmjopen-2020-040928 on 25 March 2021. Downloaded from http://bmjopen.bmj.com/ on November 7, 2023 by guest. Protected by copyright.

Table 2 Validation measures of transgender computable phenotype (CP) with provider-reported transgender status measures, in COAST HIV-positive cohort

CP	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	Kappa (95% CI)
≥1 transgender ICD (ever)	27.5 (17.8 to 39.8)	99.7 (99.6 to 99.8)	40.4 (26.7 to 55.7)	0.3 (0.2 to 0.4)
≥1 transgender ICD (recent)	8.7 (3.6 to 18.6)	100.0 (99.9 to 100.0)	85.7 (42.0 to 99.4)	0.2 (0.1 to 0.3)
≥1 transgender ICD and ≥1 hormone/blocker use (ever)	27.5 (17.8 to 39.8)	99.8 (99.6 to 99.8)	43.2 (28.7 to 58.9)	0.3 (0.2 to 0.5)
≥1 transgender ICD and ≥1 hormone/blocker use (recent)	7.3 (2.7 to 16.8)	100.0 (99.9 to 100.0)	83.3 (36.5 to 99.1)	0.1 (0.0 to 0.2)

COAST, Comparative Outcomes and Service Utilization Trends; ICD, International Classification of Disease; PPV, Positive Predictive Value.

Concordance

There was high concordance between transgender-specific diagnoses and exogenous sex hormone or androgen blocker prescription use in this study. That nearly half of those with at least one transgender-specific diagnosis had been dispensed hormones or blockers in the past year is consistent with findings from the USA and Canadian studies (48.9% and 43.0%, respectively)[20 32 33]—suggesting face validity for the current CP.

CP development and validation

The best-performing CP overall successfully identified cisgender people who were truly cisgender (specificity) and correctly identified transgender people who were truly transgender (0.2% false-positive rate, results not shown). However, the selected CP had relatively low sensitivity, missing approximately 72.5% of 'true' transgender people in COAST, as identified by the gold standard provider-based measure. Though a relatively small proportion of the 'true' transgender sample was identified in this study, the impact on future analyses comparing health outcomes for transgender and cisgender groups is likely negligible, as even the large proportion of 'true' transgender people misclassified as cisgender (approximate n=496) is a very small proportion of the total COAST sample. At worst, this misclassification would bias results related to disparities between transgender and cisgender health toward the null, producing a conservative attenuated effect in COAST and other such administrative datasets. Further, as discussed below, gender identity classification will likely greatly improve as transgender care shifts further into the fee-for-service system in BC. As in other Canadian administrative data studies, low sensitivity may be explained in part by provider and system billing preferences using three-digit ICD diagnosis coding instead of the more specific four-digit coding and inconsistencies in the BC billing management system.[37] Despite the low sensitivity, CP development in this study with high specificity offers an advancement for transgender health research. A measure that correctly identifies cases for transgender samples in research with good success translates to better opportunities to include transgender people in health studies and to investigate their health relative to other groups. While future research may lead to improvements in CP development, the CP identified in the current study with good specificity, although relatively poor sensitivity, has important utility in advancing opportunities in transgender health research. Additionally, while differential follow-up time can affect algorithm performance, the similar mean and range follow-up time for all CPs in this study suggests that differential follow-up time was not an important source of bias in this study.

The limited agreement between the CP and provider-reported transgender status may be due to the widely

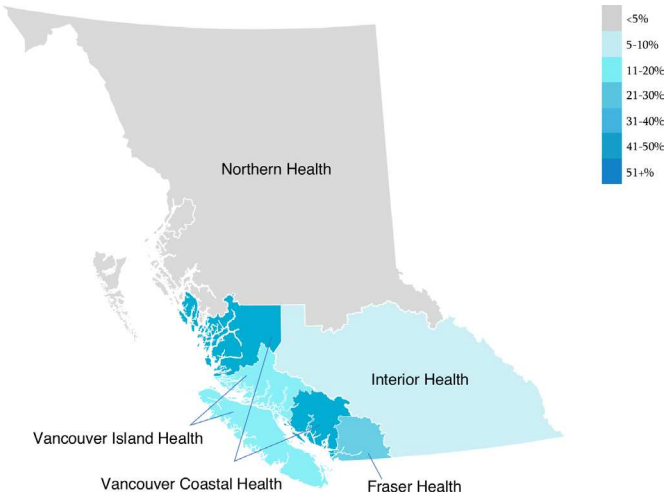

Figure 2 Geographic distribution of transgender people across province, by health authority. % of transgender individuals, among those with known health authority (n=182).

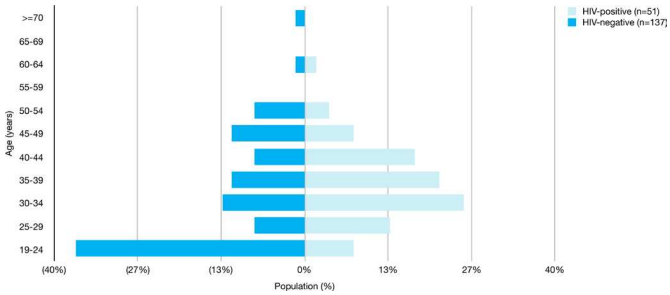

Figure 3 Age distribution of transgender sample by HIV serostatus.

BMJ Open: first published as 10.1136/bmjopen-2020-040928 on 25 March 2021. Downloaded from http://bmjopen.bmj.com/ on November 7, 2023 by guest. Protected by copyright.

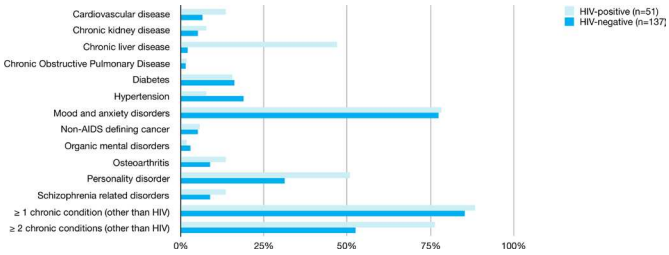

Figure 4 Comorbidities among transgender sample by HIV serostatus.

varying transgender status prevalence depending on study design and ascertainment measures used.[14] In the BC context, the CP and the DTP measures are assessing transgender status in different ways and for different purposes. In the DTP, transgender status is ascertained in the context of HIV diagnosis and ART prescribing, during which demographics and HIV transmission risk factors are recorded. This differs from recording diagnoses in EHR for those accessing transgender-specific care as used in the CP. This may explain the lower Positive Predictive Value (PPV) for the best-performing CP compared with the CP based on recent transgender diagnoses, suggesting the DTP provider-reported transgender status measure has better coverage for recent cases and the potential for use of recent diagnosis over ever to be beneficial in future CP development. Ultimately, a single CP may not be sufficient for all intended purposes and the best applicable CP (using different types of diagnoses, prescriptions or procedures) may differ depending on the intended healthcare, health research or health policy application.[17]

There is limited literature on EHR-based studies with the ability to validate an administrative transgender measure using a 'gold standard' comparison measure.[16] The two previous studies that have developed and validated algorithms to identify transgender individuals have both been conducted in non-representative samples in the USA, one using Medicare data[38] and one in a university medical centre.[16] Similar to the current study, the Medicare study found high specificity when comparing an EHR-based and a two-step survey-based transgender measure. However, the Medicare study found that the EHR measure performed consistently well with high sensitivity and a high Kappa statistic, unlike in the current study. Using chart review as the 'gold standard' for comparison of transgender status, Ehrenfeld *et al* found a low false-positive rate for their best-performing algorithm (3%), though not as low as the false-positive rate in the current study. The overall high levels of agreement for transgender measures in the two previous studies is likely a function of the lack of independence between the 'gold standard' and the CP or algorithm measures. Specifically, only those classified as transgender in the Medicare EHR data were offered survey participation to complete the two-step 'gold standard' survey measure, and only those cases identified as transgender in the university clinic EHR were included in chart review. Thus, previous studies could assess agreement between the two measures, but not robustly validate either. In the current study, the DTP provider-based transgender status measure is independent and thus could be used for robust CP validation.

While not possible to incorporate free-text records in case-finding algorithms in the current study as only structured EHR data is linked through COAST, it is worth noting the opportunities potentiated by use of Natural Language Processing (NLP) and machine-learning approaches as methods for identifying transgender samples in EHR data as this research area continues to grow. Outside of transgender health, the use of NLP and machine learning to mine unstructured free-text EHR data has demonstrated efficiency in improving case ascertainment algorithm accuracy .[39] As 'gold standard' two-step sex assigned at birth and current gender identity measures of transgender status[12] are slowly being implemented in routinely collected healthcare data sources, in the meantime NLPs to extract free-text data can be used to produce better gold standards against which to measure algorithm performance, as demonstrated by the Medicare study.[38]

Transgender status prevalence and ascertainment

Based on a recent meta-analysis of transgender status prevalence in population-based probability samples,[10] it was expected that an effective CP would identify 0.4% of the general population as transgender, or approximately 54 of the HIV-positive COAST cohort (n=13 907) and 3098 of the HIV-negative cohort (n=516 340). Consistent with expectations, the best-performing CP identified 51 PLWH as transgender, equivalent to a transgender status prevalence of 0.4% among PLWH. Contrary to expectations, the best-performing CP identified less than 5% of the number of transgender persons expected in the HIV-negative cohort. This is likely a result of a number of factors including the limitation of CPs to the subset of a population accessing care as noted, and the result of most transgender people in BC receiving care currently outside the main fee-for-service healthcare delivery system. However, it is also consistent with the undercount of transgender populations using diagnostic criteria compared with other methods of ascertainment demonstrated in other studies.[14]

Using the broadest CP algorithm (any transgender-specific diagnosis ever, n=56) and those identified by provider-reported together (total n=106), the total transgender PLWH sample would represent as high as 0.88% (range: 0.73%–1.1%) of the prevalent HIV infections in BC in 2014.[40] This over-representation of transgender people among PLWH is consistent with evidence of a disproportionate HIV burden for transgender populations globally,[5 41 42] as well as in line with the only other available data on the proportion of PLWH who are transgender, from US national surveillance data (2012 data: 1.1%, 95% CI: 0.8% to 1.4%).[43]

BMJ Open: first published as 10.1136/bmjopen-2020-040928 on 25 March 2021. Downloaded from http://bmjopen.bmj.com/ on November 7, 2023 by guest. Protected by copyright.

Demographics and chronic conditions

Despite moderate to low performance by some validation metrics, particularly low sensitivity, the CP was able to detect meaningful results in the characterisation of demographics and chronic condition burden for the transgender sample—supporting CP face validity. The population density and age distribution by HIV-status of transgender people in this study is largely consistent with general population patterns, as well as the larger COAST cohort.[21 36] The overall higher burden of chronic illness for transgender people living with HIV versus without HIV in this study is consistent with elevated chronic illness risk and morbidity among non-transgender PLWH.[44] This higher chronic disease burden is linked to HIV disease processes and related inflammatory immune response.[45] While a small but growing number of studies have begun to investigate the chronic illness burden for transgender populations in other industrialised settings,[16 19 46–48] including using EHR data, findings vary widely due to differences in sampling, study design, setting and measurement.

Limitations

Findings from this study should be interpreted in the context of a few key limitations. CPs are by design only applicable to people accessing healthcare services, often motivated by illness and aided by the ability to access care. As such, this study is limited to those transgender people accessing medical transition care in BC and may only represent 24%–47% of the total transgender population.[33] This study was also limited by the inability to validate the transgender CP among the HIV-negative COAST cohort, as a 'gold standard' provider-based transgender measure was only available for the HIV-positive cohort. It is possible that the transgender CPs would perform differently in populations living without HIV, particularly as healthcare contact is higher among populations living with HIV. Additionally, this study should be considered in light of the context in which it was conducted, an environment in which transgender healthcare delivery in BC is currently shifting from specialised care settings to the main primary care fee-for-service settings. Given that COAST only includes fee-for-service data, this study was limited by the inability to capture transgender people who access transgender care outside the fee-for-service system. However, fortunately, as the shift to the fee-for-service system occurs, transgender ascertainment via CPs in BC will likely improve. The administrative data used in this study may also be susceptible to coding error (and coding biases/practices) across conditions and settings,[49] potentially introducing misclassification bias in terms of transgender ascertainment. Finally, chronic condition prevalence data reported in this study should be interpreted with caution, given potential selection bias by serostatus in the COAST cohort; though any such bias likely resulted in conservative estimates of difference by serostatus in this analysis.

CONCLUSION

This study makes a number of important contributions to the literature on innovative methods in transgender health. Major contributions include development and validation of a transgender CP, using a population-based representative source population, in the Canadian context. Another strength is the approximately complete ascertainment of the population of transgender PLWH in BC, and capacity to estimate transgender status prevalence among PLWH. In a current funding environment of limited support for longitudinal transgender health studies in the USA and none to date in Canada, this study and the methods employed offer an efficient, replicable and cost-effective way forward in creating electronic cohorts for advancing transgender health research.[15] Moreover, the recent rollback of sexual orientation and gender identity data collection and legal changes in insurance coverage of transgender healthcare in the US potentiate decline in accurate claims coding for gender-affirming care.[30] This highlights the utility of work in this area from other jurisdictions, particularly those with transgender-inclusive universal healthcare systems such as Canada.

Future research should build on the methods developed in this study and explore complimentary approaches for gender identity ascertainment in administrative and EHR data, such as machine-learning approaches, as have been used to develop algorithms based on healthcare utilisation data in other research areas. Finally, the current study lays the foundation for future work with the ability to study transgender health and healthcare use patterns over time, with linkage to laboratory data, as well as inclusion of appropriate comparison groups.[15 50]

Author affiliations
[1]School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada
[2]British Columbia Centre for Excellence in HIV/AIDS, Vancouver, British Columbia, Canada
[3]Department of Social Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[4]Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada

Twitter Ashleigh J Rich @thatajrich

Acknowledgements The authors thank the COAST study participants, the BC Centre for Excellence in HIV/AIDS, the BC Ministry of Health, BC Vital Statistics Agency, PharmaNet and the institutional data stewards for granting access to the data and Population Data BC for facilitating the data linkage process. In addition, we would like to thank the COAST core team members and other support staff at these institutions for their administrative assistance with the data access and preparation.

Contributors AJR led the study from conceptualisation to analysis plan to interpretation, drafting of the first manuscript version, revisions and final version. RH acquired study data and funding. TP, MK, PS, TS and RH all contributed to study design, interpretation of results and reviewed manuscript versions. JL and MY contributed to study analysis and reviewed manuscript versions. All authors provided critical review of first and subsequent manuscript drafts, approved the final version and agree to be accountable for the work presented.

Funding This work was supported by the Canadian Institutes of Health Research, through an Operation Grant (grant number 130419), a Foundation Award to RSH (grant number 143342) and a Doctoral Research Award to AJR (grant number

BMJ Open: first published as 10.1136/bmjopen-2020-040928 on 25 March 2021. Downloaded from http://bmjopen.bmj.com/ on November 7, 2023 by guest. Protected by copyright.

152382) and support from the British Columbia Centre for Excellence in HIV/AIDS. The DTP receives funding from the provincial government of British Columbia (PharmaCare).

Disclaimer The funders had no role in the study design, analysis, interpretation of the data, drafting of the manuscript or in the decision to submit for publication.

Map disclaimer The depiction of boundaries on the map(s) in this article does not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. The map(s) are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. The data used for this study are held by the BC Centre for Excellence in HIV/AIDS under the authority of the BC Ministry of Health; as they contain confidential patient health records including HIV serostatus, data cannot be made available to other parties.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Ashleigh J Rich http://orcid.org/0000-0003-3742-7086
Travis Salway http://orcid.org/0000-0002-5699-5444

REFERENCES

1 Bauer GR. *Making sure everyone counts: considerations for inclusion, identification, and analysis of transgender and transsexual participants in health surveys. What differ sex Gend make Gend sex health Res casebook*. Vancouver, Canada: Canadian Institutes of Health Research (CIHR), Institute for Gender and Health (IGH), 2012: 59–67.
2 Bauer GR, Hammond R, Travers R, *et al*. "I don't think this is theoretical; this is our lives": how erasure impacts health care for transgender people. *J Assoc Nurses AIDS Care* 2009;20:348–61.
3 Poteat T, German D, Flynn C. The conflation of gender and sex: gaps and opportunities in HIV data among transgender women and MSM. *Glob Public Health* 2016;11:835–48.
4 Wanta JW, Unger CA. Review of the transgender literature: where do we go from here? *Transgend Health* 2017;2:119–28.
5 Herbst JH, Jacobs ED, Finlayson TJ, *et al*. Estimating HIV prevalence and risk behaviors of transgender persons in the United States: a systematic review. *AIDS Behav* 2008;12:1–17.
6 Becasen JS, Denard CL, Mullins MM, *et al*. Estimating the prevalence of HIV and sexual behaviors among the US transgender population: a systematic review and meta-analysis, 2006-2017. *Am J Public Health* 2019;109:e1–8.
7 Marshall Z, Welch V, Thomas J, *et al*. Documenting research with transgender and gender diverse people: protocol for an evidence map and thematic analysis. *Syst Rev* 2017;6:35.
8 Reisner SL, Poteat T, Keatley J, *et al*. Global health burden and needs of transgender populations: a review. *Lancet* 2016;388:412–36.
9 Cathcart-Rake EJ, Lightner DJ, Quevedo FJ, *et al*. Cancer in transgender patients: one case in 385,820 is indicative of a paucity of data. *J Oncol Pract* 2018;14:JOP2017027714:270–2.
10 Meerwijk EL, Sevelius JM. Transgender population size in the United States: a meta-regression of population-based probability samples. *Am J Public Health* 2017;107:e1–8.
11 Flores AR, Herman JL, Gates GJ. *How Many Adults Identify as Transgender in the United States? [Internet]*. Los Angeles, CA: Williams Institute, 2016. https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf
12 The GenIUSS Group. *Best practices for asking questions to identify transgender and other gender minority Respondents on population-based surveys*. Los Angeles, CA: The Williams Institute, 2014.
13 Bauer GR, Braimoh J, Scheim AI, *et al*. Transgender-inclusive measures of sex/gender for population surveys: mixed-methods evaluation and recommendations. *PLoS One* 2017;12:e0178043.
14 Collin L, Reisner SL, Tangpricha V, *et al*. Prevalence of Transgender Depends on the "Case" Definition: A Systematic Review. *J Sex Med* 2016;13:613–26.
15 Reisner SL, Deutsch MB, Bhasin S, *et al*. Advancing methods for US transgender health research. *Curr Opin Endocrinol Diabetes Obes* 2016;23:198–207.
16 Ehrenfeld JM, Gottlieb KG, Beach LB, *et al*. Development of a natural language processing algorithm to identify and evaluate transgender patients in electronic health record systems. *Ethn Dis* 2019;29:441–50.
17 Richesson R, Smerek M, Rusincovitch SUhlenbrauck G, ed. *Electronic health records-based phenotyping*. Bethesda, Maryland: NIH Collab Living Textb Pragmatic Clin Trials, 2014.
18 Quinn VP, Nash R, Hunkeler E, *et al*. Cohort profile: study of transition, outcomes and gender (strong) to assess health status of transgender people. *BMJ Open* 2017;7:e018121.
19 Blosnich JR, Brown GR, Shipherd Phd JC, *et al*. Prevalence of gender identity disorder and suicide risk among transgender veterans utilizing Veterans health administration care. *Am J Public Health* 2013;103:e27–32.
20 Proctor K, Haffer SC, Ewald E, *et al*. Identifying the transgender population in the Medicare program. *Transgend Health* 2016;1:250–65.
21 Eyawo O, Hull MW, Salters K, *et al*. Cohort profile: the comparative outcomes and service utilization trends (coast) study among people living with and without HIV in British Columbia, Canada. *BMJ Open* 2018;8:e019115.
22 BC Centre for Excellence in HIV/AIDS. Drug Treatment Program [Internet], 2010. Available: http://www.cfenet.ubc.ca/drug-treatment-program [Accessed c].
23 Population Data BC. About PopData [Internet], 2017. Available: https://www.popdata.bc.ca/aboutus
24 British Columbia Ministry of Health, Population Data BC. Medical Services Plan (MSP) Payment Information File [Internet], 2014. Available: http://www.popdata.bc.ca/data
25 British Columbia Ministry of Health, Population Data BC. Consolidation File (MSP Registration & Premium Billing) [Internet], 2014. Available: http://www.popdata.bc.ca/data
26 Canadian Institute for Health Information, Population Data BC. Discharge Abstract Database (Hospital Separations) [Internet], 2014. Available: http://www.popdata.bc.ca/data
27 British Columbia Ministry of Health, Populatrion Data BC. PharmaCare [Internet], 2014. Available: http://www.popdata.bc.ca/data
28 British Columbia Ministry of Health, Population Data BC. PharmaNet [Internet], 2014. Available: http://www.popdata.bc.ca/data
29 British Columbia Vital Statistics Agency, Population Data BC. Vital Statistics Deaths [Internet], 2014. Available: http://www.popdata.bc.ca
30 Arnold JD, Nelson AE, Loubier EM. Trends in insurance coverage for Gender-Affirming surgeries. *JAMA Surg* 2018;153:972.
31 Radix A, Sevelius J, Deutsch MB. Transgender women, hormonal therapy and HIV treatment: a comprehensive review of the literature and recommendations for best practices. *J Int AIDS Soc* 2016;19:20810.
32 Rotondi NK, Bauer GR, Scanlon K, *et al*. Nonprescribed hormone use and self-performed surgeries: "do-it-yourself" transitions in transgender communities in Ontario, Canada. *Am J Public Health* 2013;103:1830–6.
33 Scheim AI, Bauer GR. Sex and gender diversity among transgender persons in Ontario, Canada: results from a respondent-driven sampling survey. *J Sex Res* 2015;52:1–14.
34 Chronic Disease Information Working Group. *BC Chronic Disease and Selected Procedure Case Definitions [Internet]*. British Columbia, Canada: BC Ministry of Health, 2019: 1–132. https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-access/chronic-disease-registries-case-definitions.pdf
35 Population Data BC. British Columbia Cancer Agency. BC Cancer Registry Data [Internet], 2014. Available: http://www.popdata.bc.ca/data

BMJ Open: first published as 10.1136/bmjopen-2020-040928 on 25 March 2021. Downloaded from http://bmjopen.bmj.com/ on November 7, 2023 by guest. Protected by copyright.

36 BC Stats. Sub-provincial Population Estimates [Internet], 2018. Available: file://localhost/Users/AJRich/Zotero/storage/5VM2LBUN/PopulationEstimates.html
37 Hu W. *Diagnostic codes in MSP claim data. program monitoring and information management branch, resource management division, medical services plan*, 1996: 39.
38 Gerth J, Becerra-Culqui T, Bradlyn A, *et al*. Agreement between medical records and self-reports: implications for transgender health research. *Rev Endocr Metab Disord* 2018;19:263–9.
39 Liao KP, Cai T, Savova GK, *et al*. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *BMJ* 2015;350:h1885.
40 BC Centre for Disease Control. *HIV in British Columbia: annual surveillance report 2015*. Vancouver, Canada: BC Centre for Disease Control, 2017.
41 Poteat T, Scheim A, Xavier J, *et al*. Global epidemiology of HIV infection and related syndemics affecting transgender people. *J Acquir Immune Defic Syndr* 2016;72(Suppl 3):S210–9.
42 Baral SD, Poteat T, Strömdahl S, *et al*. Worldwide burden of HIV in transgender women: a systematic review and meta-analysis. *Lancet Infect Dis* 2013;13:214–22.
43 Centers for Disease Control and Prevention. *Behavioral and clinical characteristics of persons receiving medical care for HIV Infection—Medical monitoring project, United States, 2012*. Washington, DC: Centers for Disease Control, 2015: 42. https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-HSSR_MMP_2012.pdf
44 Grabovac I, Veronese N, Stefanac S. Human immunodeficiency virus infection and diverse physical health outcomes: an umbrella review of meta-analyses of observational studies. *Clin Infect Dis* 2019:ciz539.
45 Nasi M, De Biasi S, Gibellini L. Ageing and inflammation in patients with HIV infection: ageing, inflammation and HIV infection. *Clin Exp Immunol* 2017;187:44–52.
46 Judge C, O'Donovan C, Callaghan G, *et al*. Gender Dysphoria-Prevalence and Co-Morbidities in an Irish Adult Population. *Front Endocrinol* 2014;5:87.
47 Cheung AS, Ooi O, Leemaqz S, *et al*. Sociodemographic and clinical characteristics of transgender adults in Australia. *Transgend Health* 2018;3:229–38.
48 Swartz JA, Ducheny K, Holloway T, *et al*. A latent class analysis of chronic health conditions among HIV-positive transgender women of color. *AIDS Behav* 2019. doi:10.1007/s10461-019-02543-3. [Epub ahead of print: 29 May 2019].
49 Fisher ES, Whaley FS, Krushat WM, *et al*. The accuracy of Medicare's Hospital claims data: progress has been made, but problems remain. *Am J Public Health* 1992;82:243–8. doi:10.2105/AJPH.82.2.243
50 Reisner SL, Radix A, Deutsch MB. Integrated and gender-affirming transgender clinical care and research. *J Acquir Immune Defic Syndr* 2016;72(Suppl 3):S235–42.

BMJ Open: first published as 10.1136/bmjopen-2020-040928 on 25 March 2021. Downloaded from http://bmjopen.bmj.com/ on November 7, 2023 by guest. Protected by copyright.